Breakthrough Collaborative's Tutoring Program: Math Knowledge Gains and Participant Math Perceptions

Breyon Williams, Connor Rooney, and Greg Chojnacki

March 2023

Contents

Figures

.

Middle Years Math Grantee Report Series

This report is one in a series of six reports on math tutoring programs. Over the 2020–2021 and 2021–2022 school years, the Bill & Melinda Gates Foundation invested in rapid-cycle evaluations of a cohort of 10 tutoring providers to learn about their innovative approaches to tutoring as part of its Middle Years Math body of work.[1] The goal of these investments was to understand how different tutoring models might create positive student experiences and lead to improved academic outcomes for students in the foundation's priority communities—those who are Black, Latino, and/or experiencing poverty. These investments were grounded in the substantial body of evidence supporting the effectiveness of tutoring in improving student math knowledge (Nickow et al., 2020).

To build on this existing evidence of effectiveness, the Gates Foundation sought to develop new early evidence about the success of a range of tutoring approaches. Specifically, these investments targeted two key learning priorities. First, the foundation sought to learn how innovative technologies and tutoring program design features might simultaneously improve the quality and lower the cost of tutoring, making high-quality tutoring available to a large number of students in priority communities. The second priority was to learn the extent to which tutoring programs resulted in positive experiences for participating students. To learn about tutoring design features, the foundation invested in tutoring programs with a wide range of approaches, including group and one-on-one tutoring, virtual and in-person models, professional teachers as tutors, or volunteer tutors who shared aspects of identity with tutored students. Tutoring programs also used different approaches to tutoring curriculum and pedagogy. The goal of this report series is to inform the tutoring field more broadly and support the provision of high-quality tutoring to as many students in the priority communities as possible.

To learn rapidly about tutoring providers' innovative approaches, Mathematica worked with each one to identify the most rigorous study design that would be feasible for district partners within a one-to-three-month planning period. Some providers were able to design and implement randomized controlled trials; others used quasi-experimental designs such as matched comparison approaches. One study compared growth in math knowledge among participants to the growth observed in national samples because it was not possible to obtain student-level data for comparison students who did not receive tutoring. These relatively small studies were right-sized to the development stage of the tutoring program and sought to demonstrate early evidence of success before moving on to larger-scale effectiveness studies. To help synthesize findings about student experiences from multiple providers, studies used the same student survey measures of tutor relationship, math confidence, and sense of belonging in tutoring sessions. Most of the studies used standardized math knowledge assessments aligned with Common Core State Standards.

Each study also aimed to inform providers' efforts to refine their programs and support successful implementation. These studies measured the amount of tutoring offered, attendance, and staff impressions about implementation challenges while also gathering qualitative data on students' experiences. Findings from these studies have helped to direct tutoring providers' next steps in refining and scaling their tutoring programs.

[1] This publication is based on research funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Executive Summary

What is the tutoring program we studied?

Breakthrough Collaborative operates as two dozen community-based affiliates supported by a national office. Affiliates deliver tutoring directly to students and the Breakthrough National Office provides program standards and resources, including curricula, evaluation tools, and tutor recruitment. The Breakthrough National Office alongside two of its 24 affiliates (Breakthrough Central Texas and Breakthrough Greater Boston)—together, referred to as Breakthrough— designed a math-focused, school-year version of its summer program to pilot during the 2021–2022 school year . The program (also referred to as the Breakthrough tutoring program or the tutoring program) was designed to offer high-dosage math tutoring and social-emotional learning opportunities—both incorporating culturally responsive education practices—to 8th-grade students. Breakthrough tutors are near peers; share students' racial, ethnic, and cultural background, allowing for culturally mindful mentorship; receive training on culturally responsive education, and are paid. Breakthrough planned to offer in-person tutoring to students twice per week over 21 weeks after school, with sessions lasting 120 to 135 minutes, and a mix of two-to-one and four-to-one tutor-to-student pairings. See Appendix A for additional details about the tutoring program.

What questions does this study answer?

1. Among students who participate in the tutoring program, what is the average attendance rate during tutoring sessions?
2. Do students who participate in the tutoring program report having a high-quality relationship with their tutor after receiving tutoring?
3. Do students who participate in the tutoring program report high levels of sense of belonging after receiving tutoring?
4. Do students who participate in the tutoring program report higher levels of math confidence after receiving tutoring than before the tutoring began?
5. Do students who participate in the tutoring program demonstrate gains on the district administered i-Ready assessment (Breakthrough Central Texas) or Renaissance Star Math assessment (Breakthrough Greater Boston)?

How was the study conducted?

Study design. Both affiliates offered the program after school. The Central Texas program was offered twice weekly for two hours each session. All 8th-grade students who were existing Breakthrough students with this affiliate were invited to participate. Of these 50 students, 49 students enrolled. Three sites (or campuses) served students at Central Texas. The Greater Boston program was initially offered twice weekly for three hours total but changed to once weekly following winter break. Among existing Breakthrough 8th-grade students with this affiliate, 40 students were invited to participate, and 32 students enrolled. One site served all students at Greater Boston. Students enrolled in tutoring in the Central Texas program attended sessions at their schools and students enrolled in tutoring in the Greater Boston program were transported from their school to the affiliate site for tutoring.

Program sessions consisted of a variety of activities, such as instruction, homework support, and social activities.

Measures and analysis. The findings in this report are drawn from student attendance records, survey data, math assessment data collected from districts, and student and tutor focus groups. Attendance data were collected for all program participants, but all other data were collected only for program participants who consented to the study. Tutors recorded student attendance during each tutoring session. Affiliate sites administered the Copilot-Elevate tutor caring subscale, the Copilot-Elevate classroom belonging subscale, and the Patterns of Adaptive Learning scale surveys at baseline, end of program, or both to measure the quality of students' relationships with their math tutors, their sense of belonging, and their math confidence. The i-Ready Math assessment at Central Texas and the Renaissance Star Math assessment at Greater Boston were administered by students' schools, and the data were provided to Breakthrough. Breakthrough administered two end-of-year focus groups in total: one for students that was split into two sessions and one for tutors. For attendance and survey data, the study team analyzed descriptive statistics. To learn about changes in math knowledge among participants, the study team estimated average differences between pre- and post-program test scores and compared those changes to average changes in a nationwide sample prior to the pandemic (the latest year available). For focus group data, we noted themes that emerged from student and tutor feedback to further contextualize key findings. See Appendix B for additional details about the methods.

Limitations. Low survey response rates (ranging from approximately 37 to 68 percent, depending on the specific analysis sample) affected the extent to which the sample of students included in the analysis represent all students who participated in the tutoring program and consented to participate in the study. Additionally, the use of national data that came from pre-pandemic years might have affected the accuracy of our estimates of the extent to which students' test scores would have increased, in the absence of tutoring, as we discuss in the key finding on math knowledge below. Lastly, there are several reasons why any comparisons between affiliates should be interpreted with a great deal of caution. The two affiliates served populations of students who differed on several dimensions. For example, among students with a math score record in our data, the percentage of participants who were female was 59.3 percent at Central Texas and 35.7 percent at Greater Boston. Also, the percentage of participants who were Hispanic was 81.5 percent at Central Texas and 50 percent at Greater Boston. Further, variation exists in the fidelity of implementation and contextual factors between affiliates, including how the COVID-19 pandemic affected the program in each location.

What did the study find?

Usage. Breakthrough experienced attendance challenges, although bright spots emerged. Anecdotal evidence from affiliates suggests that the COVID-19 pandemic and the competing after-school commitments of students contributed to attendance challenges. Breakthrough will continue to learn from its attendance bright spots and challenges to inform program design and implementation.

Student perceptions and beliefs. Survey responses suggest that students held favorable views about their relationships with their math tutors and about their sense of belonging in tutoring sessions, on average, at the end of programming. Survey responses also provide descriptive evidence that math confidence increased from baseline to the end of programming, although that change was not statistically significant in the small sample of survey respondents.

Math knowledge. Participants' math scores grew during the year, on average. This growth was similar to the growth from a national sample during a pre-pandemic school year. Given that the study design did not include a comparison group of similar students who did not participate in the tutoring program, it is not possible to determine conclusively the extent to which Breakthrough's tutoring program contributed to the observed growth. See Appendix B for additional details about the limitations on the approach for establishing the math knowledge finding.

Introduction

Breakthrough Collaborative operates as two dozen community-based affiliates supported by a national office. Traditionally, the Breakthrough model has consisted of two components: (1) a summer enrichment program that provides academic and social-emotional learning support to students and (2) out-of-school-time tutoring and enrichment during the school year to keep students connected between the summer months. The study focuses on a math-focused, school-year version of Breakthrough's summer program for the 2021–2022 school year. The program was designed to offer high-dosage math tutoring and social-emotional learning opportunities—both incorporating culturally responsive education practices—to 8th-grade students. Breakthrough tutors are high school and college students and share students' racial, ethnic, and cultural backgrounds. Breakthrough partnered with its Central Texas and Greater Boston affiliates to pilot the intervention. Both affiliates offered the program after school. The Central Texas program was offered twice weekly for two hours each session, and the Greater Boston program was initially offered twice weekly for three hours total but changed to once weekly following winter break.

The study aims to measure student–tutor relationships and sense of belonging among program participants. Also, the study aims to examine the relationship between participating in the tutoring program and students' math confidence and achievement. Each analysis uses a descriptive approach. In this evaluation, we sought to answer the following research questions:

1. Among students who participate in the tutoring program, what is the average attendance rate during tutoring sessions?
2. Do students who participate in the tutoring program report having a high-quality relationship with their tutor after receiving tutoring?
3. Do students who participate in the tutoring program report high levels of sense of belonging after receiving tutoring?
4. Do students who participate in the tutoring program report higher levels of math confidence after receiving tutoring than before the tutoring began?
5. Do students who participate in the tutoring program demonstrate gains on the district administered i-Ready assessment (Breakthrough Central Texas) or Renaissance Star Math assessment (Breakthrough Greater Boston)?

Key Findings

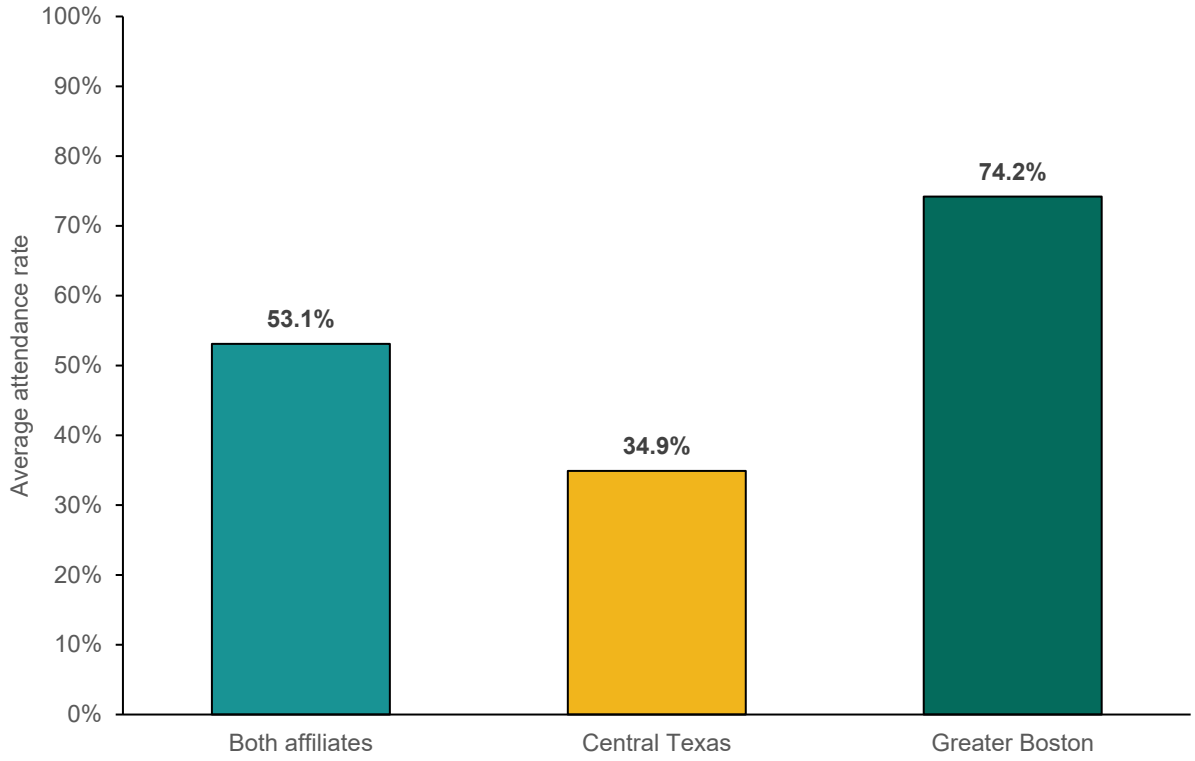
Breakthrough experienced challenges and bright spots with attendance.

Across both affiliates, students attended 53.1 percent of Breakthrough tutoring sessions, on average, throughout the school year (Figure 1). However, attendance varied across the two affiliates. Among the 49 students enrolled at Central Texas, 75.5 percent (37 students) attended at least one tutoring session. The average attendance rate was 34.9 percent for those 37 students. Among the 32 students enrolled at Greater Boston, all attended at least one tutoring session, and the average attendance rate was 74.2 percent (Figure 1). Spring semester attendance was, on average, lower than fall semester attendance across both affiliates. In the fall semester, the average attendance rate was 64.4 percent across both affiliates. In the spring semester, the average attendance rate was 46.8 percent across both affiliates. In this report, we define program attendees as students who attended at least one Breakthrough session during the school year. These students

had some exposure to the Breakthrough tutoring program and were the students whom the affiliates considered to be program participants.

A subset of students in both affiliates consistently attended programming. At Central Texas, 16.2 percent of program attendees attended at least 75 percent of the offered sessions. At Greater Boston, 62.5 percent of program attendees attended at least 75 percent of the offered sessions.

Figure 1. Average attendance rate, overall and by affiliate



Notes: Sixty-nine students attended tutoring across both districts—37 at Central Texas and 32 at Greater Boston. The average attendance rate for each affiliate was calculated among students who attended at least one Breakthrough tutoring session. The program encompassed 35 tutoring sessions at Central Texas and 20 tutoring sessions at Greater Boston. We calculated the attendance rate for each student by dividing the total number of days present by the total number of sessions held.

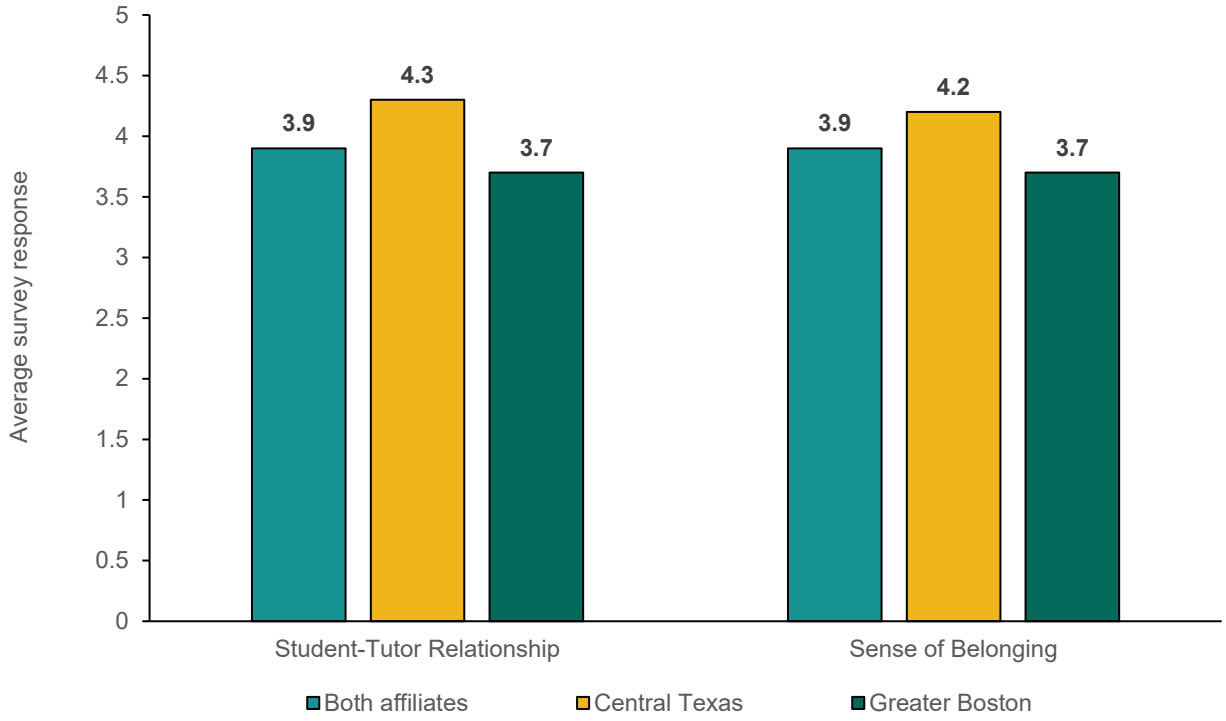
Breakthrough's focus toward the end of programming was on learning about the factors that influenced attendance; these learnings would inform design and implementation changes to boost attendance in a future iteration of the Breakthrough tutoring program. Breakthrough held one focus group with students and another with tutors to learn more about the contextual factors and student-specific challenges that may have affected attendance. Although the factors influencing attendance during the school year were multifaceted, some possible factors that emerged from the focus groups included other after-school commitments and family obligations competing for students' time, as well as the attendance of students' friends at the sessions. In addition to the learnings from the focus groups, both locations experienced COVID-related challenges that affected attendance (for example, families' decisions to keep students at home due to illness or potential exposure), though Central Texas experienced more significant disruptions to in-person programming due to the pandemic.

Students who completed the end-of-year survey reported favorable views about their relationship with their math tutors and their sense of belonging in their math tutoring sessions, on average.

Across both affiliates, the 29 students who completed the end-of-year survey (42 percent of program attendees) rated their relationship with their math tutors and their sense of belonging in their math tutoring sessions a 3.9 out of 5, on average, on a 5-point scale (Figure 2). The student–tutor relationship survey items ask students the extent to which they agree with positive statements about their relationship with their math tutors, and the sense of belonging survey items ask the same about students' sense of belonging in their math tutoring sessions. Response options range from strongly disagree (1) to strongly agree (5), where a higher score translates to a student having a more favorable view of their relationship with their math tutor or their sense of belonging in their math tutoring sessions. At Central Texas, students rated their relationship with their math tutors a 4.3 out of 5, on average, and rated their sense of belonging in their math tutoring sessions a 4.2 out of 5, on average. At Greater Boston, students rated their relationship with their math tutors and their sense of belonging in their math tutoring sessions a 3.7 out of 5, on average.

Eighty-six percent of students who completed the end-of-year survey attended at least half of all tutoring sessions, and 65 percent attended at least three-quarters of all tutoring sessions. Because most of the students who completed the end-of-year survey are students who were considered frequent program attendees, it is likely that responses reflect the perspectives of students who viewed the program more favorably and therefore participated often.

Figure 2. Average end-of-year ratings for student–tutor relationship and sense of belonging, by affiliate



Notes: Twenty-nine students across both affiliates had a response on the end-of-year student–tutor relationship construct and the end-of-year sense of belonging construct, including ten students at Central Texas and nineteen students at Greater Boston.

Math confidence increased slightly, on average, among students who completed both the baseline and end-of-year surveys, although this difference was not statistically significant.

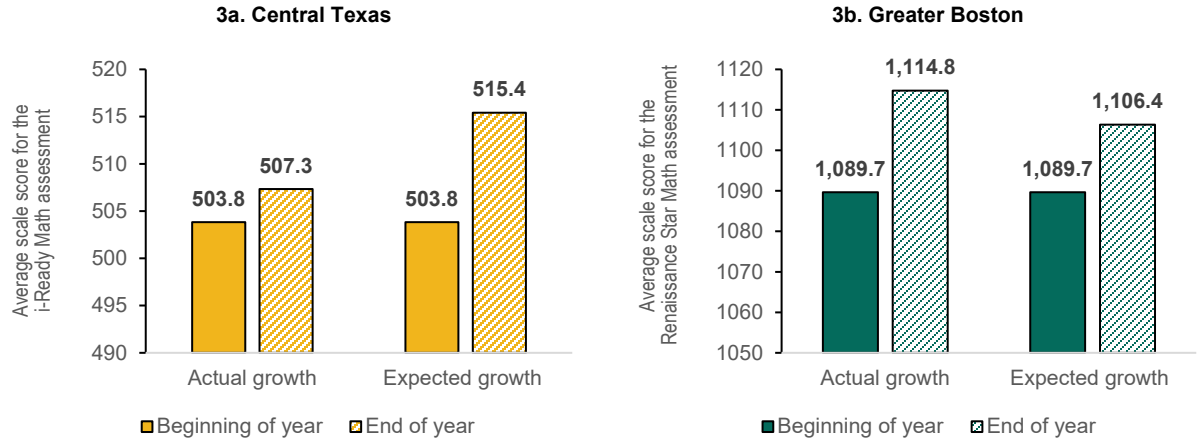
Across both affiliates, students who completed both the baseline and end-of-year surveys (21.7 percent of program attendees) rated their math confidence a 3.1 out of 5, on average, on the baseline survey and a 3.2 out of 5, on average, on the end-of-year survey, although this change was not statistically significant. The math confidence survey items ask students how confident they are in completing math tasks, and response options range from not at all (1) to very (5), where a higher score translates to a student having more math confidence. At Central Texas, the average response grew from 2.8 at baseline to 3.3 at the end of the year. At Greater Boston, the average response declined slightly—3.2 at baseline and 3.1 at the end of the year.

Participating students' math scores grew from beginning to end of the year across both affiliates, on average, and these gains were similar to gains from a national sample in a pre-pandemic school year.

Among the small sample of 32 Breakthrough participants whose math assessment scores were measured by the study, scores increased, on average, from the beginning to the end of the 2021–2022 school year at Central Texas and Greater Boston. At Central Texas, students' scale scores on the i-Ready Math assessment increased 3.5 points, on average (Figure 3). We refer to this as the Central Texas students' *actual growth* in 2021–2022. Still, the actual growth might have happened even if students had not participated in the tutoring program, because Breakthrough students were also receiving regular math instruction in school over the same period. We used national score data from the most recently available school year, 2018–2019, to estimate what growth would be for Central Texas students in a normal school year for students with the same fall 2021 score, without the tutoring program (11.6 scale score points, on average). We refer to this as students' *expected growth* in the absence of the tutoring program. Then, we compared students' actual growth in 2021–2022 against their expected growth, absent the tutoring program. The difference between actual and expected growth was –8.1 scale score points, on average. Considering this difference as a share of the spring 2022 test's standard deviation (34 points), the difference translates to an effect size of –0.24 standard deviations.

At Greater Boston, actual growth in scores on the Renaissance Star Math test was larger than expected growth, based on national data from school year 2016–2017, and the corresponding effect size was 0.12 standard deviations, using the test's standard deviation (72 points) from the most recently available national data. Across both affiliates, the average effect size was –0.02 standard deviations, meaning that actual growth was similar to expected growth in the absence of the tutoring program, on average. Neither the affiliate-specific estimates nor the combined estimate was statistically significant due to our small sample of students with baseline and outcome test scores.

Figure 3. Actual and expected growth on math tests, Central Texas and Greater Boston



Notes: Figures are scaled so that the vertical axis for each figure has a range of approximately 1 standard deviation on the test. See Appendix B for more details on assessments. Twelve students at Central Texas and 20 students at Greater Boston have both beginning-of-year and end-of-year math scores. The i-Ready Math test score range is 0 to 800, and the Renaissance Star Math score range is 0 to 1,400.

Although comparing participants' actual growth against expected growth based on these national samples helps to contextualize the change in test scores that participating students experienced, it is important to interpret this comparison with caution. Specifically, if students generally experienced less growth in math knowledge during the 2021–2022 school year than in pre-pandemic years, then our comparison understates the degree to which these Breakthrough participants' scores may have increased, relative to expected growth. In addition, for both affiliates, the data used as a benchmark for participants' actual growth is from a national sample of test takers. This means that the data may not provide the most accurate indication of expected growth for Breakthrough participants because these students likely differ from the national sample. See Appendix B for additional details on analysis methods and related considerations when interpreting these findings.

Anecdotal data yielded some insights on facets of implementation that might have limited gains in math knowledge for Breakthrough participants, particularly at Central Texas.

- **Time spent on homework instead of instruction.** At the request of students, some tutoring sessions featured an increased amount of tutoring time allocated to homework support rather than math instruction, although data gathered by Breakthrough does not detail the extent of this shift. Breakthrough also reported that students needed time not only to complete math homework but also homework in other subjects, potentially diverting additional time from math instruction.
- **Attendance challenges.** The attendance findings show that many students did not receive a full or close-to-full dosage of tutoring, and this could have limited the extent to which the tutoring program affected math knowledge. As noted above, the average attendance rate and effect size on math knowledge were larger at Greater Boston, suggesting that dosage matters.

Participants in the student and tutor focus groups expressed the importance of having more flexibility in how students are supported during the tutoring sessions.

In the focus groups, students and tutors reported desiring more flexibility with the tutoring schedule across tutoring sessions. In an interim implementation assessment, Breakthrough observed that

some students, particularly in Greater Boston, expressed concern that they needed to get homework done during the program because it would be late in the evening when they returned home. To address these concerns, Breakthrough conducted two focus groups (one for students and one for tutors) to determine whether changes to the tutoring schedule are needed to improve the design of its tutoring program.

Breakthrough found that students' and tutors' preferences about the share of time allocated to instruction versus homework support varied. Students who participated in the focus group felt the tutoring was useful but reported that they had limited time to complete their homework after school because of a series of after-school obligations. Students expressed a desire to have the option to choose when instruction was provided and when homework support was provided, which might include reserving time for students to complete their homework. Tutors who participated in the focus group expressed different opinions about how much time should be allocated to homework support (ranging from at least one-fourth of the time to most of the time).

Autonomy among Breakthrough affiliates and their tutors has traditionally been an essential component of Breakthrough's program, ensuring that affiliates are responsive to the specific needs of the community they serve. Because this pilot was focused primarily on math tutoring, a subset of the support that Breakthrough affiliates typically offer, the affiliates exercised somewhat less autonomy than they normally would.

Overarching Conclusion and Next Steps

Students who experienced the Breakthrough tutoring program established positive relationships with their math tutors and felt a strong sense of belonging in their math tutoring sessions. Across the two affiliates, students' math knowledge also grew. On average, this growth was similar to the growth from a national sample in a pre-pandemic school year. Challenges in data collection and student attendance ultimately prevented the study from reaching strong conclusions about the effectiveness of the Breakthrough tutoring program. However, one clear takeaway is that given the prominence of attendance challenges in Breakthrough tutoring this past school year, Breakthrough believes that future research on strategies to improve attendance would benefit the field.

Breakthrough identified several potential changes to the program design to address the key findings identified in this report:

- **Establishing school-level collaborations to improve the attendance and the usefulness of the tutoring program.** Breakthrough observed potential benefits associated with establishing relationships with school faculty and staff. First, teachers might limit the amount of homework assigned to students on tutoring days. Second, math teachers might provide tutors with diagnostic information on students to help tutors personalize the tutoring experience. Third, Breakthrough could coordinate with teachers and other school staff to identify support services for students with attendance or other challenges. Although school-level collaborations hold promise for improving the Breakthrough tutoring program, Breakthrough must weigh the benefits against the potential for increased burden on school partners to maximize buy-in from school staff.
- **Scheduling students' tutoring on days with minimal competing after-school commitments to improve attendance.** Feedback from end-of-year surveys revealed that students may have forgone tutoring because of competing after-school commitments. Gathering information from

students on their commitments before setting tutoring schedules could reduce the barriers students face to attending tutoring.

- **Using a different curriculum that better supports Breakthrough's efforts to deliver culturally responsive education practices.** For this grant, Breakthrough was required to use a curriculum that EdReports had rated as meeting expectations for alignment to college and career readiness standards and for usability. Breakthrough was unable to find one that it felt sufficiently incorporated culturally responsive education practices. For example, Breakthrough was interested in using problem-solving approaches that would have been culturally relevant and engaging (from resources such as Math Circles and Citizen Math, although these resources are not curricula) but felt its ability to do so was limited. Academic rigor and cultural responsiveness are related and important factors that affect the outcomes examined under this grant. Breakthrough believes that one role an afterschool tutoring program can play is to provide alternative ways for students to access core concepts and learning opportunities in areas where the core curriculum used in class does not fully meet students' learning needs. This is particularly important when the core curriculum does not include as many engaging and culturally inclusive presentations of concepts and practice opportunities. It is possible that the efficacy of the tutoring program, particularly on math confidence and math knowledge, may have been greater if the curricula used for tutoring offered more flexibility.

References

Bruch, J., Forde, J., Lai, I., and Sotelo Muñoz, V. (2022). Menu of measures: High-quality measures of Middle Years Math student outcomes. Mathematica. https://www.mathematica.org/-/media/publications/pdfs/education/2022/menu-of-measures_math_march-2022.pdf

Curriculum Associates, LLC. (2021). i-Ready diagnostic national norms for reading and mathematics for grades K-8. https://www.fldoe.org/core/fileparse.php/7539/urlt/iready-norms-tables-K-8-2020.pdf

Nickow, A. Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on preK-12 learning: A systematic review and meta-analysis of the experimental evidence. National Bureau of Economic Research. https://www.nber.org/papers/w27476

Swain, M., Randel, B., & Dvorak, R. N. (2019). Impact evaluation of mathematics i-Ready instruction for middle school grades using 2018-19 data (Final Report No. 109). Human Resources Research Organization. https://eric.ed.gov/?id=ED604747

What Works Clearinghouse. (2022). What Works Clearinghouse standards handbook, version 5.0. Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/handbooks

Appendix A. Detailed description of tutoring program

Breakthrough designed a math-focused, school-year version of its summer program for the 2021–2022 school year. The tutoring program was designed to provide rigorous, culturally relevant instruction to current 8th-grade Breakthrough students from two affiliates (Breakthrough Central Texas and Breakthrough Greater Boston). The program also consisted of other activities outside of instruction, including homework support and social activities.

Key components of the program

- Breakthrough is intentional about serving students from historically marginalized communities. In 2020, 96 percent of students participating in the summer Breakthrough program identified as a person of color, 77 percent received free or reduced-price lunch, and 75 percent would be the first in their family to attend college.
- Breakthrough tutors receive training and ongoing professional development support from experienced instructional coaches. Breakthrough prioritizes recruiting tutors who are high school and college students from underrepresented groups in the teaching pipeline (STEM majors, male students, people of color), former and current teaching fellows, former Breakthrough students, people with experience or interest in teaching or studying math or STEM, and those with a passion for social justice.
- Students receive up to 44 hours of math instruction aligned to Breakthrough culture and state standards. Instruction is meant to be culturally responsive and rigorous and takes place over no more than 21 weeks in a mix of 4:1 and 2:1 small-group tutoring sessions at Central Texas and Greater Boston. In addition, students are offered nonacademic activities outside of regular math instruction aligned to the Breakthrough culture.

Appendix B. Methods

Design. The study team analyzed the attendance records, data on survey and math knowledge, and qualitative data from focus groups using a descriptive approach.

Measures. The study team collected daily attendance records on 37 students at Central Texas and 32 students at Greater Boston to measure program participation and used the Copilot-Elevate tutor caring subscale (student–tutor relationship construct), Copilot-Elevate classroom belonging subscale (sense of belonging construct), and Patterns of Adaptive Learning scale (PALS) (math confidence construct) to measure the main survey constructs. Details on each of these survey measures are available in Mathematica's menu of high-quality middle years math student outcome measures, which were selected in consultation with external measurement experts (Bruch et al. 2022). These three constructs were included in a survey that was administered at the end of the program, and the math confidence construct was also included in a survey that was administered at the beginning of the program. Among enrolled students, 27 percent at Central Texas and 59.4 percent at Greater Boston completed the three constructs at the end of the program. Further, among enrolled students, 16.2 percent at Central Texas and 28.1 percent at Greater Boston completed the math confidence construct at the beginning and end of the program. Central Texas administered the i-Ready Math assessment and Greater Boston administered the Renaissance Star Math assessment to measure math knowledge. Each assessment was administered at the beginning of the program and at the end of the program. Among enrolled students, 32.4 percent at Central Texas and 65.6 percent at Greater Boston had beginning-of-year and end-of-year math assessment scores. Each affiliate conducted a student focus group at the end of the program (two focus groups in total). The focus groups consisted of four students at Central Texas and six students at Greater Boston. Central Texas conducted a tutor focus group with five tutors at the end of the program.

Analysis. The study team reported average student attendance rates across both affiliates and by affiliate. Each student's attendance rate is calculated as the ratio of the number of days present to the number of total days tutoring was offered. For survey data, average responses are reported by construct and survey time point (baseline and end of year). The study team calculated, for each student, the mean response across items within a given construct. Next, the average of the means was reported across all students. Analysis of change in math confidence only included students who had survey responses at both time points. Cronbach's alpha was calculated for end-of-program surveys by affiliate for the math confidence construct. The reliability coefficient (alpha) for PALS was 0.85 at Central Texas and 0.92 at Greater Boston, both above the threshold for acceptable reliability of 0.6 adopted by the U.S. Department of Education What Works Clearinghouse (What Works Clearinghouse, 2022).

To assess whether the actual growth in math scores that students experienced while participating in Breakthrough tutoring was substantially different from what would be expected during a school year without tutoring, we used national data to estimate expected growth. Specifically, for each student, the study team identified the percentile rank associated with each student's observed baseline score, then used technical documentation from each math assessment to identify the end-of-year score the student would be expected to receive if the student remained at the same percentile at the end of the year. (The relevant technical documentation for the i-Ready was published by Curriculum Associates, LLC [2021], and the documentation for the Renaissance Star Math assessment was provided to us directly by the publisher.) The team matched students' beginning-of-year percentile rankings with end-of-year scale scores using aggregated data on a national sample of test takers for

both assessments to determine expected end-of-year scale scores. Even though individual students may score in higher or lower percentiles, on average, we would expect students in the sample to remain at the same percentile except to the extent that they or their school environments differed from the national norming sample in a way that affected average math knowledge gains. One interpretation of this is that if a student remains at the same percentile in the spring as where they were in the fall, this indicates the student experienced the same gain in math knowledge throughout the course of the year as other students who started at the same percentile and remained at the same percentile.

Although comparing actual growth to expected growth based on a national sample helps to contextualize actual growth, there are important limitations to this comparison. First, students at Central Texas and Greater Boston, or their schools, may differ from the national sample in important ways that would make their expected growth in the absence of tutoring lower or higher than expected growth for a national sample. Second, the national sample data for both tests were drawn from a pre-pandemic year, when students may have experienced different levels of growth, on average, than they did in school year 2021–2022. Specifically, 2016–2017 was the most recent school year for which national data on expected growth were available for the Renaissance Star Math assessment, and 2018–2019 was the most recent available year of data for the i-Ready assessment. If average growth in those pre-pandemic years was higher than average growth in the 2021–2022 school year, then the expected scores we used as a benchmark are artificially high and would overstate the degree to which we expect Breakthrough participants' scores to rise in the absence of the tutoring program. Third, when translating differences between actual and expected average scale scores into effect sizes, we divided those differences by the standard deviation of end-of-year scores in each assessment's national norming sample. The standard deviation for the Renaissance Star Math assessment in spring 2022 for a national sample of 8th-grade test takers (72.2) was provided to us directly by the publisher, but the standard deviation used for i-Ready scores (33.81) was from a national sample of test takers in the 2018–2019 school year (Swain et al. 2019). If scores were more spread out during the study school year, then the overall effect size and the effect size for Central Texas that are reported would overstate the size of the differences between actual and expected scores, and the true effect sizes would be closer to 0.